

What is claimed is:

1. A method for creating a cleansed output file from an input file, comprising the steps of:

- (a) selecting an input file containing a plurality of data records;
- (b) selecting a reference files, said reference file containing a plurality of data records;
- (c) computing a search key; and
- (d) for each said data record in said input file:
 - (i) retrieving said data record from said input file on remote storage;
 - (ii) searching said reference file with a matcher process for all said data records in said reference file that match said search key and reading each said data record from said reference file that matches said search key, thereby generating a candidate data record list;
 - (iii) searching said candidate data record list and determining a matching data record, wherein said matching data record matches said data record in said input file;
 - (iv) creating a new cleansed data record;
 - (v) cleansing said data record of said input file according to said matching data record, thereby generating verified information;
 - (vi) writing said verified information into said new cleansed data record; and
 - (vii) writing said new cleansed data record to a cleansed output file;

wherein said steps (d)(i) through (d)(vii) are performed in a single pass through said data records of said input file and in a single pass through said reference file, such that each data record of said input file is read from a remote storage location only once, each said matching data record of said reference file is read from a remote storage location only once, and each said new data record to said cleansed output file is written to a remote storage location only once.

2. The method according to claim 1, further comprising two or more reference files and further comprising the step of:

- (e) repeating said step (d)(v) for said data record of said input file wherein said matcher cleanses said data record of said input file using said two or more reference files.
3. The method according to claim 1, further comprising two or more reference files and two or more matcher processes, wherein each said matcher process access one of said one or more reference files and said data record of said input file resides in local memory while being processed by each of said matcher processes and each of said reference files.
 4. The method according to claim 3, further comprising the step of:

(e) recycling one said data record of said input file through one or more matcher processes while said data record of said input file resides in local memory, wherein said recycling processes said data record of said input file through one or more of said one or more reference files previously accessed.
 5. The method according to claim 3, further comprising the steps of:

(e) each said matcher process determining whether a subsequent matcher process should process said data record of said input file, wherein said determining results in no additional processing is performed on said data record of said input file, one or more said matcher processes are skipped, or a subsequent matcher process is changed.
 6. The method according to claim 3, further comprising two or more search keys wherein each said matcher process accesses one said search key.
 7. The method according to claim 2, further comprising two or more search keys, wherein said matcher process performs step (d)(ii) using each of said search keys.
 8. The method according to claim 1, wherein said search key comprises: a predefined number of digits of the USPS ZIP+4 Code, a predefined number of digits representing an address number of a postal patron, a predefined number of alphanumeric characters representing a street name of the postal patron, and a predefined number of alphabetic characters representing the postal patron.
 9. The method according to claim 1, wherein said reference file is organized such that two or more data records of said reference file that match a predefined search key are stored on remote storage in a primary physical block of memory, such that said matcher process reads said candidate data record list in said step (d)(ii) with a single memory read command.

10. The method according to claim 9, wherein said reference file is organized such that two or more data records of said reference file that match a said search key are stored on remote storage in a primary physical block and an overflow physical block of memory, such that said matcher process reads said candidate data record list in said step (d)(ii) with two memory read commands.
11. The method according to claim 1, wherein said determining said matching record by said matcher process of said step (d)(iii), comprises the steps of:
 - (1) for each data field in said search key, determining a degree of match between said data record of said input file and a data record of said reference file in said candidate data record list, said degree of similarity ranging in value from an identical value to no identifiable similarity;
 - (2) arranging said degree of match determined in said step (d)(iii)(1) in a table;
 - (3) determining whether each row of table in said step (d)(iii)(2) represents a match or no match of said data record of said input file and said data record of said candidate data record list; and
 - (4) establishing for each row of said table in said step (d)(iii)(2) a match or no-match value.
12. The method according to claim 1, wherein said data records of said input file contain a mixture of personal and business data records, said business records containing zero or more contact names and addresses, and said search key containing data elements selected from the group consisting of personal name and demographical information.
13. The method according to claim 1, wherein in said step (d)(iii) said matcher process matches a female's maiden name to her married name, comprising the steps of:
 - (1) matching a confirming piece of information, such as all or part of the: Social Security Number (SSN), Date of Birth (DOB), Driver's License State and Number, Phone Number, USPS Delivery Point Code, or other identifying information of the female in said matching data record and said data record of said input file, or matching a first name and at least a middle initial in said matching data record and said data record of said input file;
 - (2) matching a portion of a last name in said matching data record and said data record of said input file; and

- (3) confirming that the gender of the female is selected from the group consisting of female, indeterminate, or blank.
14. The method according to claim 1, further comprising the steps of:
- (e) generating a unique front door key for a front door of an address wherein said unique front door key comprises a USPS ZIP+4 Code + a predefined number of digits of a house number and a predefined number of characters of a unit number;
 - (f) assigning said unique front door key to said cleansed data record of said output file; and
 - (g) maintaining a reference database that correlates all new DPC assignments for an address to said unique front door key.
15. A method for assigning and maintaining a unique and unchanging unique Person Key using a multitude of available but not consistently available data across all such person's absolute and indicative keys and descriptive and associated data, comprising the steps of:
- (a) maintaining a base record including not necessarily all of the persons parsed first name, nick name, middle name, last name, including hyphenated last names, NYSIIS or SOUNDEX or some similar encoding of first name and of last name, generational suffix, gender, SSN, DOB, Driver's License State Abbreviation and Number, up to five current or prior Phone Number's, last five USPS Delivery Point Codes (DPC) or parts thereof, and an assigned unique key which shall be called the Person Key;
 - (b) maintaining records from all available sources shall be processed such that if a minimum acceptable name (such as, the person's first initial and last name) matches with at least one of: (1) so or all Digits of the SNN (SSNn), (2) at least DOB CCYYMM, (3) the Driver's License, or (4) other identifying number, the records will be called a match and all descriptive information combined into a single record;
 - (c) matching records from all available sources such if a good name match is made (such as first name, middle initial if present, and last name, and generational suffix if present and is confirmed by one or more of Phone Number, DPC, and any other identifying information that may be available in two or more data file sources, the records will be called a match and all descriptive information combined into a single record;
 - (d) matching records from all available sources such if a complete name match is made (such as, first name, middle name, last name and if present in one record generational suffix must be present in all and must agree), the

records will be called a match and all descriptive information combined into a single record;

(e) indexing one or more of the following:

- i. Person Key;
- ii. Clear text version, NYSIIS or SOUNDEX of Last Name+All or part of SSN;
- iii. Clear text version, NYSIIS or SOUNDEX of Last Name+ All or part of DOB;
- iv. All or part of SSN+All or part of DOB;
- v. First Name+All or Part of SSN;
- vi. Phone Number;
- vii. Driver's License Number;
- viii. Surname+First Initial+Middle Initial;
- ix. DPC;
- x. Other versions of the information in the name and descriptive information file;

(f) retrieving on any of the available keys using information from an input record and then to match, using an available matching tool (such as weights-and-penalties matcher or table driven matcher) that input record to the base record to then assign a person key; and

(g) optionally appending a footnote code to each input record that describes the degree of match to each data field of an input record.

16. The method according to claim 1, wherein the method provides data to a third party with associated software to allow use of the provided data, without allowing the third party direct access to the data in clear text form, wherein said reference database is un-encrypted and said method further comprises:

(e) a means for encrypting said reference file and extracting a key to allow aggregating records to be considered for match between said input file and said reference database;

(f) a means for encrypting said input file and extracting said key for said reference database;

(g) a means for comparing said encrypted input file against said encrypted reference file, and if there is an acceptable match in encrypted form to cause said encrypted reference data record to be unencrypted, and the required data from said reference file be appended to said data record of said input file; and

(h) reporting the data content used for royalty reporting purposes.

17. The method according to claim 16, wherein the method prevents the third party which has been provided said encrypted reference file and associated software for use of said encrypted reference file from modifying usage reporting of said encrypted reference file without detection of such modifications.

18. The method of claim 1, wherein said method provides USPS postal coding without depending on the ZIP Code or Last Line City and State Abbreviation to be correct, comprising the steps of:

(e) wherein said reference file contains USPS ZIP+4 data, indexing said reference file on:

(i) Delivery Point Code (DPC);

(ii) Five Digit ZIP Code+Right most N (such as five) characters of House Number+First M (such as three) characters of street name;

(iii) House number+SOUNDEX street name (without prefixes and suffixes)+unit number (if any)+Street PreDirectional (if any)+Street Post Directional (if any)+Street Type (if any); and

(iv) State+City Name;

wherein said matcher process comprises a means to retrieve on any of the available keys using information from an input record and then to match, using an available matching tool that input record to the base record to then assign the postal codes and parsed address text, and a means for appending a footnote code to each input record that describes the degree of match to each data field of the input record.

19. The method according to claim 3, wherein the method improves match rates for historic and “dirty data” address files, wherein one said reference file is constructed from the Federal Information Processing Standards (FIPS) Named Populated Places file, said search key indexes on place name and state, and said matcher process returns a ZIP Code from such index Named Populated Places reference file, further comprising:

(e) matching using this ZIP where there was no ZIP Code present in the input record or the ZIP Code was different for the associated place name, or no match was obtained with the input record provided ZIP Code; comprising;

- i. inputting the original input ZIP Code to obtain a current ZIP Code;
- ii. inputting the original state and place name to obtain the current ZIP Code; or

- iii. proceeding with a matching process using the new ZIP Code obtained;
- (f) creating a second reference file constructed from the long term history of discontinued ZIPs and associated discontinued place names and the new ZIP Code and new place name;
- (g) indexing said second reference file on the state and place name and also on ZIP Code;
- (h) proceeding, if no match was found with the matching process using either of the following:
 - (i) inputting the original input ZIP Code to obtain a current ZIP Code, or
 - (ii) inputting the original state and place name to obtain the current ZIP Code; and
- (i) proceeding with a matching process using the new ZIP Code obtained.